DARPA MIND'S EYE PROGRAM

Frequently Asked Questions

DARPA-BAA-10-53

April 29, 2010

Defense Advanced Research Projects Agency
3701 North Fairfax Drive
Arlington, VA 22203-1714

Frequently Asked Questions

Q1.	Can you elaborate on specific obligations of those pursuing the systems integration tasks with respect to transitioning visual intelligence technology?
A1.	Systems integrators will be expected to develop a visual intelligence subsystem composed of at minimum a processor and camera, that manifests the visual intelligence capabilities provided by visual intelligence team, meets power and weight requirements outlined in the BAA and that meets the interface and API specifications that will be provided by DARPA's evaluation and transition partner, Army Research Lab (ARL) for a target robotic platform. Whether the subsystem makes use of power and communication systems afforded by the robotic platform or integrates is own will be at the discretion of the systems integrator but must comply either way with the indicated specifications. In other words, the degree to which the subsystem is self-contained or depends on the target platform is discretionary.
Q2.	Do you expect research in visual intelligence will result in one best approach?
A2.	Not necessarily. We are open to pursuing multiple approaches as long as each is showing improvements and relevance to the program goals.
Q3.	Will systems integrators be obligated to implement every approach coming out of Phase I? Or, do you anticipate that systems-integrator/visual-intelligence-developer pairs to arise "organically" through the PI meetings? Do the responses to these issues change if a team is doing both visual intelligence development and system integration (presuming the developed visual intelligence approach is a success in Phase I)?
A3.	Systems integrators will not be obligated to implement every visual intelligence approach. Pairings will occur organically and perhaps with some facilitation by DARPA. There is no guarantee that a team offering both visual intelligence and system integration proposals will be (a) awarded both contracts, (b) kept together throughout the program, (c) advanced at the same pace for both efforts. Furthermore, there are no principled restrictions on components of a combined team being combined with other visual intelligence developers and/or system integrators.
Q4.	If systems integrators are expected to re-implement multiple visual intelligence approaches, will DARPA enforce or institute any commonalities to reduce development costs and risk (e.g. common infrastructure such as SDKs, operating system, or APIs)?
A4.	While it is anticipated that some commonalities may exist across approaches and hardware, DARPA wishes to avoid any specifications that would rule out unorthodox technical approaches. Thus, if in their perusal of candidate visual intelligence partners, systems integrators identify facilitative constructs, it will be up to them to convince candidate partners to adopt those constructs. It will be further up to the visual intelligence teams to advocate on their own behalf in preservation of their

	technical innovations.
Q5.	Do you anticipate specific measures to protect both visual intelligence developers' and system integrators' intellectual property? E.g., will licensing issues be up to the performers, or do you anticipate a standard licensing agreement?
A5.	The Mind's Eye program seeks to advance fundamental research in this nascent space. Thus, it is hoped that proposers will apply due consideration to the program objective of seeding an emergent research community with algorithms, methods, and corpora to support ongoing related work. In fact, this objective is described in the evaluation criteria section of the BAA. DARPA will not dictate the licensing terms but leave open the possibility of requesting changes after the proposals are received. Please also keep in mind the implicit notion of "government use" as defined in the Federal Acquisition Regulations (FAR and DFARS).
Q6.	In the BAA there is a link to the DARPA TCTO website, where both teaming information and videos will be posted. Do you know when those links will be made available on the DARPA website?
A6.	There is a "coming soon" notice on the Mind's Eye program website. Some files will be made available on April 15th, but others may follow, so stay tuned.
Q7.	Under Section E - Funding Restrictions, it is stated that DARPA currently anticipates using 6.2 funding for this program. Has a final determination yet been made?
A7.	Mind's Eye is a 6.2 funded program.
Q8.	Are proposers allowed to submit separate proposals for both technical areas? Conversely, is it permissible to submit one proposal that covers both technical areas?
A8.	Proposers who are interested in pursuing both technical areas are requested to submit a single proposal with a separate technical section and costing for each area. To quote the BAA, page 12, "Proposers may offer solutions for both technical areas, and in this case may be selected to perform either or both activities. Proposers who expect to contribute in both areas can expect visual intelligence development and systems integration concepts to be independently evaluated. In this case additional systems integration costs should be carefully justified."
Q9.	Is it necessary to have an industry partner?
A9.	No. Any proposer to the system integration effort should demonstrate relevant integration knowledge and experience as indicated in the BAA.
Q10.	Is there an expectation that the majority of support will go to University performers?
A10.	There is no expectation either way. Academic and industrial organizations are encouraged equally to submit proposals.
Q11.	Page 6. Does the recognition of primitive actions include recognizing the arguments of those actions? Example: "Approaches" presumably takes two arguments as in Approaches(moving object, place)

	"Tom approached the checkpoint" Approaches(moving object, moving object) "Tom approached Bill as he was coming up the path" If so, this could require recognizing a large collection of objects including balls, vehicles, people, weapons, boxes, checkpoints, buildings, etc. Will such a recognizer be GFE? If not, will training data be labeled with these object classes? Or is each team expected to provide or acquire this capability?
A11.	The expectation is that teams will employ state-of-the-art computer vision capabilities as necessary to support their respective technical approaches. Such capabilities will not be provided by DARPA in order to avoid constraining technical approaches to a particular representation. While one technical approach may require such a compositional breakdown in the input, another approach may be more holistic at the input level.
Q12.	Page 7. "apply spatiotemporal events learned in one setting to another (e.g., from indoor office scene to outdoor tactical environment)." Will this capability be evaluated? Changes in lighting, object types, background objects can be severe in going from indoor to outdoor scenes.
A12.	Over the course of the program, and with advance notice, the development and evaluation corpora will evolve to incorporate increasing complexity. Dimensions of complexity will be elaborated over the course of the program but will be geared toward developing a capability that is general and can be operationalized in a real-world environment. Thus, early corpora will be simple, with constraints such as consistent lighting, simple textures, ample pixels on target, and minimal occlusion. In addition to constraints related to canonical computer vision, there will also be early constraints on complexity as it applies to visual intelligence (e.g., the use of simple vs. composite events, where a composite event such as “replace” is composed of “take” followed by “put”). Over time, both computer vision and visual intelligence constraints will be relaxed. The sample vignettes available in advance of the industry day exemplify some representative points on the complexity continuum. However, DARPA is sensitive to the notion that the current state of the art in computer vision (object recognition, segmentation, etc.) has limitation, and seeks to focus the efforts of this program on advancing visual intelligence rather than traditional computer vision. Thus, *DARPA is receptive to ideas about the appropriate character of the complexity trajectory in support of advancing the program’s research goals, which are welcome in proposals*.
Q13.	Page 7. "...over time develops a sense for what is ordinary and out-of-the-ordinary through continued exposure to a particular visual scene." This requires normalcy models and/or anomaly detection methods. Will this be evaluated? What kind of visual scenes do you have in mind?
A13.	In learning by discovery, one can imagine that over the course of experience clusters of similar patterns might aggregate. Such clusters could indicate classes of events. In the context of such perceptual space, exemplars that fall outside of these

	classes could be construed as anomalies. This isn't to suggest a particular method of detecting anomalies, but rather to exemplify the notion of anomaly in this context. It is anticipated that such a capability would be integral to discrimination among events. DARPA does not rule out the possibility of evaluating such a capability either directly or indirectly toward advancing research goals and operational capabilities. Examples can be dangerous in that they might lead to narrow expectations. However, the following example is provided with the caveat that it is a single point in a potentially very large space of possibilities: "Consider a bus stop on a country road. A commonly observed event at this bus stop is people *exiting* and *entering* a bus and *sitting* on a bench. An anomalous event would be a person *leaving* the bus stop on foot."
Q14.	Page 7 and elsewhere. "It does this through direct examination". What alternative to direct examination are you trying to exclude by this statement?
A14.	The idea of "direct examination" is emblematic of what human's call "envisionment", that is, the notion that one can examine a spatio-temporal representation rather than some symbolic manifestation of that in order to benefit from the uniqueness and expressivity of the features of that representation. For example, if one envisions a person holding ball, it is easy to answer the question "is the person touching the ball" by directly examining the representation. Answering that question based upon a propositional representation of a person holding a ball (e.g., isHolding(person,ball)) would require a supportive ontology that contains knowledge about how hold relates to touch, and how touch relates to contact. That being said, there's the notion of parsimony in symbology. Thus, just as humans build an ontology based on experience to support future reasoning, one might expect a visual intelligence system to also properly use perception to build or augment an ontology and then to subsequently employ that ontology to support future reasoning. But in any case, the direct examination of the spatiotemporal representation was the vehicle by which the knowledge entered into the ontology and, hence, a precursor to reasoning symbolically.
Q15.	Page 7. Issuing alerts. The F1 metric (page 20) treats false alarms and missed alarms as equally bad. Is this the right tradeoff? One might instead employ a metric that bounds false alarms such as true detection rate subject to false alarm rate < X%. If the goal is to maximize the number of events that a fixed set of analysts can handle, then perhaps that should be the metric (number of events successfully detected for a given amount of analyst time).
A15.	It is important to keep distinct the notions of evaluating the performance characteristics of the visual intelligence system under development and flexibility in the fielded system. The idea espoused on page 20 in the BAA addresses the former and is simply the notion that information retrieval metrics will likely be appropriate to evaluating performance in tasks for which there are likely to be multiple correct responses and there is interest in assessing a practicable balance of the completeness (avoiding misses) and precision (avoiding false alarms) in that set.

	DARPA is making no early commitments to the exact form of this metric, which may even evaluate precision and completeness separately. In order to address the notion of flexibility in the fielded system, one might imagine a "bias parameter" in algorithms that would permit real-time adjustment of the trade-off between precision and completeness.
Q16.	Page 7 and elsewhere. How will interpolation and prediction be evaluated?
A16.	The representative evaluation rubric for integrated cognition on page 20 addresses this. The specifics of this rubric are evolving and have not yet been articulated in detail for that reason. To stimulate thinking, though, one might imagine a scene in which two people approach each other from opposite directions and pass behind a structure, which occludes them. One person enters carrying a briefcase. When they emerge the other person has the briefcase. This scene may be presented to the system with a query: "what happened behind the structure?" As a benchmark, human responses may be aggregated to form a gold standard set to these queries. Information retrieval theory might then be applied to the analysis. But, to reiterate, this is still evolving.
Q17.	Page 20. "IED emplacement" is a complex activity made out of several verbs. Are activities of such complexity seriously being considered? Are activities that the actors are trying to hide being considered?
A17.	The pursuit of composite event recognition (events composed of more than one atomic event in one or more related sequences) is within the program's scope. The complexity of composite events (i.e., activities) that are in scope, however, will depend upon a variety of factors including (but not limited to) technical progress. Surreptitious activity is also within scope in that it calls for spatiotemporal gap analysis, as articulated in the BAA. However, also in this case, target complexity will be governed by a variety of factors.
Q18.	Page 9. "the boy approaching the soldier intends to give the item to the soldier". This requires goal/intent recognition. Will it be sufficient simply to predict that the next event is going to be Give(Boy, Item, Soldier)? Note similar "intent" statements on Page 15 under "Grounding".
A18.	Yes, prediction will suffice. However, one can imagine that goal/intent recognition may not be so far out of reach in the context of outcome prediction (as already posited in the program) and the notion that some behaviors are goal-directed and some aren't. If intent classification can be applied to the latter, interesting possibilities arise. Perhaps intent classification lends itself entirely to spatio-temporal analysis in certain contexts. The exploration of such possibilities is not precluded from proposals – in fact, such innovations that are consistent with a technical approach already leading to core capabilities are welcome.
Q19.	Page 15. "Visual Inspection". We found this description very confusing. Is the following interpretation correct?

	Given a video sequence, the system will recognize and label the sequence (or objects/events in the sequence) using a hierarchy of relationships ranging from simple spatial relationships (distance(A,B)=0.5m, A touches B) to force dynamic relationships (A supports B, A presses against B) to verbs (A carries B) to vignettes (A picked up items B1,B2,B3 and delivered them to persons P1, P2, and P3). Labels will apply to image regions or segments.
A19.	The emphasis here is not on a specific technical approach, but rather that some representations may lend themselves to spatiotemporal analysis in support of recognition and reasoning better than others. In particular, it is the qualitative (i.e., relational and dynamic) aspects of the scene that seem more relevant to this kind of analysis than compositional characteristics of objects. So notions like spatial relationships, motion vectors, and perhaps higher moments are all consistent with this idea, though that list is by no means exhaustive.
Q20.	Page 15. "Grounding". Is the following interpretation correct? The system designer (or perhaps the user) should be able to provide declarative knowledge to the system. This knowledge should be (a) refined through experience and (b) combined with knowledge learned from video analysis (spatio-temporal patterns, prediction and interpolation rules, etc.) during all aspects of reasoning (e.g., recognition, prediction, interpolation, question answering).
A20.	This interpretation seems essentially correct, though perhaps instead of saying "during all aspects of reasoning" it would be more accurate to say "as supported by visual learning and opportunistically in support of reasoning".
Q21.	Page 19. What is the definition of an "unfamiliar event"? Page 16. What is the definition of "new, surprising inputs"? For example, will the system be required to do additional learning (e.g., of novel verbs) during the evaluation? Will novel objects or object types be arguments to known verbs?
A21.	Unfamiliar in this context means "previously unseen". "Event" connotes one instantiate of an event class. A specific example of an event class is "give", which includes all instances "give", reflecting a rich variety of manifestations and contexts. Thus, an "unfamiliar event" is not suggestive of a new event class, but rather a new instantiation of a previously learned class, like "give". So if the system learned "give" from watching a boy give a ball to a man, it would be expected to recognize an unfamiliar event, such as a boy giving food to a dog. However, to ensure generality in learning and not just recognition, it is anticipated that learning new event classes will also be part of the evaluation. Thus, the notion of "new, surprising inputs" may be construed to suggest both new "events" to be recognized and new "event classes" to be learned dynamically and then generalized.
Q22.	Page 19. What is the definition of "learn by discovery"?
A22.	To learn by discovery is to acquire the ability to classify events in the absence of supervision. For example, over the course of learning, a perceptual space may

	emerge naturally that results in clusters of event instances in the space. If the system can partition the space on the basis of those clusters, then new exemplars could be classified on the basis of those class boundaries. The notion of classifying events is independent of assigning English labels to event classes.
Q23.	Page 14. Can you give an example of a visual concept or spatio-temporal relationship that could be learned by envisionment?
A23.	If the following query were posed: "Does give require approach?", the system could envision "give" in its prototypical forms or instantiate various manifestations of give to support or refute that correlation.
Q24.	Page 14. "Proposers should describe the learning mechanics in detail, how users would interact with a deployed system" This could refer either to user interactions to teach the deployed system new verbs and new scenarios. Or it could simply involve prioritizing alerts for previously-trained scenarios, obtaining envisionments, etc. Which is intended?
A24.	The quoted text refers to the former, that is, teaching the deployed system operationally relevant concepts, or perhaps setting parameters that support discovery-based learning that is consistent with user goals and interests. Generally, the idea is to describe how a user would train a deployed system. However, insofar as the training methodology has implications for interactions related to operational use, it is worthwhile to describe those considerations as well.
Q25.	Page 14. Envisionment could have many different purposes: (a) Explanation of a recognized event or scenario. The system could overlay envisionment as a way of highlighting the recognized events. This could also support debugging and learning. (b) Summarization of recognized events/scenarios. The envisionment could take the form of a compact summary (perhaps even just some key frames). (c) Debugging of new event definitions (not corresponding to any observed videos). The user could enter a symbolic event description and request an envisionment of it. This could be useful for debugging such event descriptions. (d) Generating synthetic training data. As in (c), the user could ask the system to synthesize video and then feed it to the lower level learning components to train tracking, object recognition, verb recognition, etc. Note that (c) and especially (d) are vastly more difficult than (a) or (b), as they require learning detailed models of appearance and dynamics that, while not needed for recognition, are important for producing useful synthetic training data. How would you prioritize these with respect to the goals of the program?
A25.	(a) and (b) are a higher priority than (c) and (d) and more central to the program objectives. Among other things, the latter would seem to build upon the technical underpinnings of the former.

Q26.	The BAA states that "DARPA envisions a typical visual intelligence development team will consist of a core of 3-5 personnel." Was the intent to suggest that the team consists of 3-5 people total (i.e. a University professor or two with a handful of grad students) or 3-5 lead contributors plus developers to support each lead contributor (i.e. 3-5 professors or industry tech leads who each bring a small contingent of students/engineers)?
A26.	The BAA also explicitly mentions “additional contributors of visual input processing and symbolic reasoning components as needed to perform the fundamental research necessary for this technical area” so there is not a hard restriction with regard to team size. However, besides bringing in additional contributors for specific aspects of the work, the intent of the guidance is that the core team would be made up of 3-5 people total, and not 3-5 lead contributors supported by a small contingent of students/engineers. We would take this opportunity to point out that ultimately, proposers are welcome to propose any team composition deemed suitable for the proposal’s approach and expected results.
Q27.	It was unclear from the BAA whether proposals offering component technologies or partial solutions would be considered compliant with the BAA. Must proposals cover complete end-to-end solutions to be compliant or will proposals for partial/component solutions be accepted?
A27.	Proposals should respond to technical areas in their entirety. Thus, offerors with partial solutions should consider teaming with others who have complementary offerings. Please visit the teaming site at https://visint.org.
Q28.	Given how close the industry day is to the proposal deadline, will the government consider an extension on the proposal due date?
A28.	No. DARPA hopes that the Industry Day meeting will be helpful to proposers, but expects that the BAA along with this document will provide enough interim guidance to enable proposers to make significant headway on proposals in advance of the Industry Day. **Change 1:** Although the submission deadline will remain May 10, the deadline time is changed from 12:00 noon to 12:00 midnight following. This is to ease the difficulties that the previous time posed for those in more western time zones.
Q29.	Pg 16 of the BAA talks about previous experience in integrating perceptual systems in the "visual spectrum"...Does this imply that the choice of camera is visible and not IR?
A29.	While the integration of non-visual spectra as well as other sensory modalities is not precluded from future work, the Mind’s Eye focus will be on perception in the visual spectrum. In fact, the test stimuli used in the evaluation of systems will be limited to that spectrum, so even if a candidate system could make use of extravisual data, it wouldn’t be present in the test signal. However, if a particular technical approach benefits substantially by extending the input spectrum, it would be interesting to know about that, even if it weren’t central to the proposal.
Q30.	The BAA allocates 10 pages to section 2.4 technical approach and 10 pages to

	section 2.7 Statement of Work. Are these correct? If so, please provide guidance.
A30.	The page allocation for each section is an upper bound. Please use the space parsimoniously. Proposals will be evaluated on the merits of their substantive content – not on length. Also, it is always helpful to preface larger sections with summary statements.
Q31.	The BAA allocates 10 pages to section 2.4 Technical Approach. Do teams proposing both tasks (visual intelligence and system integration) have 20 pages?
A31.	Yes, if necessary but do not exceed the 10 pages restriction when describing your technical approach for either task. Please be succinct and ensure that all content contributes directly to an understanding of the technical approaches. **Change 1:** If you are submitting a proposal to address both tasks, you may also double the stated page limit for the following sections of your proposal: 2.1, 2.5, 2.7, and 2.10.
Q32.	On p. 13, in the visual learning section, the document states that "these concepts are to be learned directly from visual inputs (video), and in terms of a generally applicable representation". Later, it also states that "It is expected that a range of learning techniques may be applicable, and that these may entail varying degrees and techniques of supervision." What seems unclear is the type of allowed supervision. Does it consist just of annotations of the training data, or can the representations themselves also be specified to some degree in the training process?
A32.	There are practical reasons to prefer an approach that minimizes supervision, which is why minimal supervised learning is indicated in the Program Scope of the BAA (Page 6). Thus, any approach to supervision, particularly those that involve heavier supervision, should include an explication of the attendant cost/benefit analysis. For example, does a particular supervision method result in substantially faster learning? How robust is the system to having less supervision? How easy would it be for a non-technical user to provide necessary supervision?
Q33.	Is it expected that learning takes place from a single example or can we expect multiple examples to be available?
A33.	The evaluation data will include multiple examples of events within event classes. For example, the training input may include several examples of the event class “give”. Nonetheless, in order to better understand the performance characteristics of a system, learning may be assessed both during and after training.
Q34.	Is it possible to have a one-on-one meeting with the Program Manager at Industry Day or after?
A34.	Once the BAA is released, DARPA will not discuss potential responses or potential concepts or approaches relevant to the Mind's Eye BAA with potential BAA responders. Any discussions or responses to such conversations could be construed as advice or direction regarding BAA responses, and could provide an unfair advantage to a potential BAA proposal. Questions can be submitted to the BAA

	mailbox with any answers provided to be available to the entire community so as to prevent any unfair advantage, real or perceived, to a specific proposer.
Q35.	What role are universities best suited for: primes, subcontractors, solo proposers?
A35.	Any of those are viable options for universities. We have no guidance to give on this matter.
Q36.	Is the set of verbs given in the BAA exhaustive? Is there interest in extending the list? Might verbs which deal with human interactions be added (such as verbs describing emotion)?
A36.	As was stated during the Mind's Eye Industry Day, the set of verbs is not considered closed. Proposers are invited to identify how the proposed approach relates to this list of actions, and how one would approach that list, add to it, etc. But it is important to have a goal for the program and constraining the set of target verbs is one way to bring focus to the program's research.
Q37.	Will objects be identifiable by monochromatic uniform color (as seen in some of the sample vignettes)?
A37.	The sample corpora will offer a range of vignettes. Test data will move through a range of complexity as described on pages 17-18 of the BAA. It is a challenge for us to figure out how to build a test corpus to allow for that progression but we will gauge the level of difficulty that is appropriate based on the performer community's progress.
Q38.	You are planning on mounting "smart cameras" on moving platforms. How much camera motion should be accounted for in phase I? Can we assume no camera motion in phase I?
A38.	Accounting for camera motion is outside the scope of Mind's Eye as the program is conceptualized right now.
Q39.	Why worry about the computational constraints of the UGV? Why not transmit the video and handle it at the control center?
A39.	The transformative CONOPS imagined in Mind's Eye requires that the processing be done at the platform to enable the rapid use and dissemination of the intelligence generated.
Q40.	Can you clarify how envisionment interacts with the other aspects of Mind's Eye?
A40.	We are looking to the proposers to expand upon the ideas presented in the BAA, including topics such as how envisionment might work and how you would use it. We want to foster your research in imagination and visual manipulation by the system.
Q41.	Can you say something about hand-generated world knowledge versus learned world knowledge (i.e., ontologies)?
A41.	We have no presupposition about this, and invite proposers to address that issue if it is germane to the proposed approach.
Q42.	Could you expand on the term "integrator"?
A42.	The integrator's role is as defined in Technical Area 2 of the BAA.
Q43.	Should the smart camera meet the size, weight, and power constraints of a man-portable UGV as a near-term or long-term goal?

A43.	We expect these considerations to become germane during the course of the program when phase 2 systems integration concepts are proposed.
Q44.	The program consists of components of existing systems and new research. Some "engineering" might be required to integrate existing components with new research. Is this engineering work within the bounds of the program?
A44.	It is expected that teams would propose and perform whatever systems engineering is necessary to implement their proposed solutions.
Q45.	Can integration onto a small military UGV be included in the proposal?
A45.	Proposers will not be funded to do platform (UGV) integration. This will be separately addressed by DARPA and transition partners.
Q46.	Who conducted the seedling?
A46.	MIT CSAIL
Q47.	How much funding has DARPA allocated for this BAA?
A47.	We are not providing such information.
Q48.	Is there an Industry Day attendees list that will be made available?
A48.	We do not plan on releasing that information but we would encourage you to make use to the available teaming resource available at the following URL: https://visint.org/
Q49.	How many awards are you anticipating?
A49.	See page 21 of the Mind's Eye BAA.
Q50.	If there are multiple awards will there be a down select at some point in the program?
A50.	No "downselect" is included in the program concept at this time.
Q51.	Is the Mind's Eye program limited to land-based scenarios or could sea-based or air-based platforms be included?
A51.	Mind's Eye inputs will represent ground level scenes as might be seen from the perspective of a human or UGV.
Q52.	Why is there such a short deadline for the submission of proposals?
A52.	A 45-day submission deadline ensures the earliest possible kickoff in the current year.
Q53.	Will this project be considered as fundamental research? Will universities need publication approval?
A53.	See page 22 of the Mind's Eye BAA.
Q54.	Regarding video inputs, should we plan for 24 hour (day and night) operations and therefore the use of LWIR or low-light cameras? Multi-spectral? Other sensors?
A54.	We will be using daytime scenes only initially. Video will be the input. At this time, we will not be relying on other sensors.
Q55.	Is Mind's Eye limited to a single color camera? Is visual input limited to monocular cameras or can a stereo imaging system be used?
A55.	We are not making any definitive statements regarding the cameras that might

	emerge as part of integration concepts. However, there are no stereoscopic inputs in the sample corpora or anticipated in the development corpora.
Q56.	How do you envision the level of effort to vary during the program for phase I and phase II?
A56.	No answer at this time. This will be determined based on development during the course of the program.
Q57.	Do objects in the scenes need to be recognized or would it suffice to say "person put object on head"?
A57.	Proposers should focus on actions, as described in the BAA, even if this involves some initial limitation in the range of objects to be incorporated.
Q58.	For the first in-field demonstration in the phase plan, what is the expectation on system hardware that can be used given that the System Integration task would just be starting up?
A58.	There is no particular timing relationship implied in Figure 5 of the BAA. Furthermore, any in-field demonstrations will be planned by DARPA and ARL as opportunities to be will be pursued based on the current state of development.
Q59.	The posted video examples focus on foreground agents whereas the CONOPS might suggest distant observations of activities. Can you say something about how to recognize these?
A59.	Both near and distant scenes are potentially relevant. It is assumed that closer-in perspectives (which one might assume could be obtained via optics, even if the platform is distant) provide adequately challenging conditions to start with, and that the complexity factors provided on pages 17-18 provide ample additional challenge/variation of application.
Q60.	Have you considered multiple cameras and multiple UGVs?
A60.	Not at the moment. Proposals may put forth the benefits of such ideas if those ideas make the approach more responsive.
Q61.	The BAA states, "Specifically excluded are techniques that rely exclusively on pattern recognition techniques that aggregate visual primitives into higher-level artifacts for motion in scene" Page 14. Does this rule out object recognition, composition of scene elements, tracking based approaches?
A61.	That particular statement referred to the element of visual intelligence described in that paragraph. We make no stipulations about what state of the art recognizers, for example, that you might choose to incorporate.
Q62.	Can you say more about the 40 pound payload?
A62.	The total payload includes the camera, processor and power supply (batteries).
Q63.	On Page 14 of the BAA, in the Spatiotemporal Patterns section, should the sentence reading, "Specifically excluded are techniques that rely exclusively on pattern recognition techniques that aggregate visual primitives into higher-level artifacts for motion in scene, as well as approaches that rely on exhaustive labeled corpora of variations of action and settings." Be read as excluding deep learning?
A63.	No. The most important consideration here is that these patterns enable and support the other abilities described in Visual Intelligence.

Q64.	Are visual attention mechanisms outside the scope for this program?
A64.	No, they are not outside the scope. We do not take such mechanisms to be a solved problem. You may include such mechanisms (and whatever else) you think is relevant to creating breakthroughs in visual intelligence.
Q65.	Pages 27 and 28 give conflicting guidance regarding proposal submission. Page 27 states that T-FIMS must be used, page 28 suggests that grants.gov may be used if seeking a grant or cooperative agreement. Could you clarify which is allowable and/or preferred from DARPA's perspective?
A65.	Proposers requesting grants and cooperative agreements have the option to submit through Grants.gov; however, submission through T-FIMS is preferred.
Q66.	The initial closing deadline for proposal submission is 9AM pacific time on Monday May 10. Would it be possible to move the closing deadline time such that it is less problematic for those on the west coast?
A66.	The submission deadline date will remain May 10, however the deadline time is now officially changed from 12:00 noon eastern time to 12:00 midnight eastern time May 10.
Q67.	The BAA focuses on verbs but the samples have rudimentary objects and actions. Is there any interest in being able to handle more complex/realistic manifestations?
A67.	Yes, and in fact the test scenarios will include a range of complexity as described in the BAA, pages 17 - 18.
Q68.	Is it in the best interest of the research to require that the smart camera system fit into such a small payload?
A68.	The constraints afforded by a man-portable UGV acknowledge the method of operations anticipated for future ground forces. They also provide reasonable balance between unrealistic unconstrained conditions and the crippling constraints in more austere conditions.
Q69.	Is focus on software / in silica solutions? Or are biologically implemented approaches welcomed?
A69.	All proposals are welcomed and will be evaluated.
Q70.	Can you say more about how communications will be handled? Do developers need to be concerned with the mechanism for broadcasting information? What about content?
A70.	Developers are not responsible for broadcasting the information but Mind's Eye systems will need to communicate using text and video (as stated on Page 19 of the BAA and illustrated throughout the BAA).
Q71.	Can you say more about how one ought to proceed when proposing for both the visual intelligence effort and the system integrator effort? There are many sections other than Technical Approach where it would be useful to have more pages available in order to make our approach clear.
A71.	If you are submitting a proposal to address both the visual intelligence task and system integration task, you may double the stated page limit for the following sections of your proposal: 2.1, 2.4, 2.5, 2.7, and 2.10. Do not exceed the stated page restriction per section when describing either task. Please be succinct and ensure that all content contributes directly to an understanding of your proposal.

Q72.	Is it allowable or even encouraged, to use offline learning to develop an understanding of visual actions, action sequences, and normalcy?
A72.	There are no hard constraints on learning requirements. However, some consideration should be given to the operational use of Mind's Eye technologies.
Q73.	Should the visual intelligence system be able to distinguish similar verbs such as "give" vs. "take", "leave" vs. "go", "chase" vs. "follow", or "push" vs. "touch"?
A73.	Yes, and it is recognized that these distinctions are more challenging than those between more disparate actions. Proposers are invited to address how they will address these challenges.
Q74.	Would methods that have been recently published and are likely to succeed for the Mind's Eye tasks be considered revolutionary enough to be funded?
A74.	All proposals are welcomed and will be evaluated.
Q75.	The BAA mentions that there will be no "irrelevant" objects in the video, yet there are moving and static objects that are irrelevant or only slightly relevant to the activities in the sample videos. Could you comment on the apparent disparity between what was said and what has been shown?
A75.	Pages 17 and 18 explicates a number of complexity factors and ranges along those dimensions from "basic" to "realistic" inputs. Thus, the sample vignettes that depict some irrelevant objects are intended to illustrate inputs closer to the "realistic" end of that spectrum.
Q76.	Are trained animals (or electronically hooked up ones) out of the question?
A76.	All proposals are welcomed and will be evaluated. Also, see page 34 of the BAA for guidance concerning animal use protocols.
Q77.	What hardware can be used for the evaluation proposed techniques? Can it be GPGPU or ASIC (algorithm specific integrated circuit)?
A77.	All proposals are welcomed and will be evaluated.
Q78.	What cognitive architectures or chips being developed in other DARPA programs are suitable for Mind's Eye?
A78.	Proposers are welcome to identify and take advantage of such synergies.
Q79.	How much sophistication is expected in dialog processing, pronouns, etc?
A79.	Just enough to support the Concept of Operations (CONOPS) suggested by Table 4 in the BAA.
Q80.	BAA says we can expect 100 to 600 pixels on target – is that area or height of object?
A80.	The 100 to 600 pixels on target referred to area.
Q81.	BAA says that there will be no concurrency in actions and yet the videos show quite a bit of concurrent actions. Could you comment on the apparent disparity between what was said and what has been shown?
A81.	There is no disparity in that illustration. Pages 17 and 18 explicate a number of complexity factors and ranges along those dimensions from "basic" to "realistic" inputs. Thus, the sample vignettes that depict concurrent action are intended to illustrate inputs closer to the "realistic" end of that spectrum.

Q82.	Is the uplink bandwidth set at 640x480 / 30 FPS uncompressed video or would compression or lower frame rates be allowed?
A82.	640x480 / 30 FPS is not intended to be an uplink bandwidth constraint on communication systems, but rather a minimum requirement for spatiotemporal resolution in the input to a Visual Intelligence system.
Q83.	Are all actions between humans and other objects? Or are interactions with vehicles and animals included?
A83.	Initially, the focus will be on human-based interactions, but DARPA does not rule-out the later introduction of other actor types.
Q84.	What (or who) will dictate how much specialization to build into visual intelligence (e.g., human operators specialize by training for a task, such as check points vs. scouting)?
A84.	Proposers are invited to describe the role and extent of such specialization in enabling the capabilities defined in the BAA.
Q85.	Is DARPA more interested in research on the topic of visual intelligence or system integration?
A85.	We make no distinction with regard to interest.
Q86.	If a bidder submits a proposal for both the visual intelligence task and system integration task, is it possible that DARPA will award funding for work on only one of the proposed tasks?
A86.	Please see pages 21 and 22 of the BAA.
Q87.	Can proposers bid Technical Area 1 (Visual Intelligence) as an N month baseline followed by a M month option?
A87.	In general, options are permitted and encouraged. Please see pages 16, 19, 21, 22, 29, and 36 of the BAA.
Q88.	Should proposers who are bidding both Technical Areas 1 and 2 in a single proposal propose area 2 as an option, or should it be included in the baseline with area 1?
A88.	DARPA will accept proposals configured either way.
Q89.	Is it acceptable for one organization to submit separate proposals as prime bidder in both the visual intelligence and system integration technical areas?
A89.	There is no restriction on this in principle; however, it would be appropriate for a team to carefully consider whether this is necessary in light of the guidance in the BAA and in the FAQ on how to submit one proposal that addresses both visual intelligence and system integration tasks.
Q90.	Can you please clarify if each of the sections in Proposal Section 2 (Headings 2.1, 2.2, etc.) should each begin on a new page?
A90.	The sections in Proposal Section 2 (Headings 2.1, 2.2, etc.) do not each have to begin on a new page.
Q91.	Can you clarify what information needs to be provided in the "Official Transmittal Letter"? What should this letter say, and to whom it should be addressed.
A91.	The primary purpose of an Official Transmittal Letter is to ensure that the proposal is being submitted by a person of authority - for instance, submission made by the

	Director, Office of Sponsored Research for university submissions (and not directly from the PI's). Your letter should make clear that fact clear. You may address the letter to DARPA.
Q92.	How will a late submission be evaluated?
A92.	The potential to make awards under this BAA will depend on the quality of the proposals received and the availability of funds. This applies equally to proposals submitted after the submission deadline. While late proposals are allowed, there may not continue to be funds `available for a contract after the initial round of submissions have been evaluated.
Q93.	Does a grant fall under the umbrella of Other Transaction in Award Instrument Requested?
A93.	No, a grant does not fall under the umbrella of an Other Transaction in Award. You may find more information on page 37 of the BAA or at the following website: http://www.darpa.mil/cmo/other_trans.html.
Q94.	For the cost proposal budget (assuming the program runs for 36 months), should the budget be in three, 12 month periods or should it follow the government fiscal year. If so, would you please provide the appropriate dates to use for the government fiscal year performance periods.
A94.	As specified on page 32, Use “x months after contract award” to specify performance periods. Beyond that, it is for the proposer to decide how to best cost out the work.